INTER-GENERATIONAL PHENOTYPIC MIXING IN VIRAL EVOLUTION

Claude Loverdo[1,3] **and James O. Lloyd-Smith**[1,2]
[1] *Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA*
[2] *Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA*
[3] *E-mail: loverdo@ucla.edu*

Appendix

METHODS

We use discrete-time branching processes, where time is measured in generations, which entail the cycle from a virion infecting a cell through to the release of offspring virions. To analyze the branching processes, we use generating functions which gather the information on the probabilities $p(k)$ that a virion produces k virions for the next generation: $g(z) = \sum_{k=0}^{\infty} p(k) z^k$. Standard branching process theory implies that the probability of eventual extinction e is given by the smallest positive solution to $g(e) = e$. The probability for a viral lineage to survive the early steps and lead to emergence is $P_e = 1 - e = 1 - g(1 - P_e)$. These results can be extended to multitype branching processes, with the generating map $g_i(\vec{z}) = \sum_{k_1=0}^{\infty} ... \sum_{k_n=0}^{\infty} p_i(\vec{k}) z_1^{k_1} ... z_n^{k_n}$ where $p_i(\vec{k})$ is the probability that one virion of strain i produces a set of k_1 virions of type 1, k_2 virions of type 2, etc (Harris 1963).

A useful example is the generating function for a geometric distribution of mean N, which corresponds to the distribution of offspring virions assumed in the main text. The probability to produce k virions is $N^k/(N+1)^{k+1}$, leading to:

$$g_{geom,N}(z) = \sum_{k=0}^{\infty} \frac{N^k z^k}{(N+1)^{k+1}} = \frac{1}{1 + N(1-z)}. \quad (1)$$

In particular, if a virion leads to emergence with probability P_e, and thus its lineage goes extinct with probability $1 - P_e$, the probability of emergence considering all the virions produced by the infected cell is $1 - g_{geom,N}(1 - P_e) = NP_e/(1 + NP_e)$.

EARLY MUTATION

Let us define $g_{early}(z_1, z_2)$ as the generating function for one generation, starting from a virion of the initial strain when only early mutations can occur. When writing equations for the generating function, a virion of the mutant type corresponds to a term in z_2, and a virion of the initial type to a term in z_1. A virion of the initial strain dies without infecting a cell with probability $1 - q_1$, which leads to a term in $z_1^0 z_2^0$. With probability q_1, it successfully infects a cell, where with probability μ it produces a geometric distribution of mean N^* virions of the mutant strain (term in $1/(1 + N^*(1 - z_2))$), and with probability

$(1-\mu)$ a geometric distribution of mean N_1 virions of the initial strain (term in $1/(1+N_1(1-z_1))$). The generating function $g_{early}(z_1,z_2)$ is then:

$$g_{early}(z_1,z_2) = 1 - q_1 + q_1\left(\frac{\mu}{1+N^*(1-z_2)} + \frac{1-\mu}{1+N_1(1-z_1)}\right). \quad (2)$$

With s_2 the survival probability of a lineage initiated by one virion of strain 2, the probability of emergence P_e is solution of $1-P_e = g_{early}(1-P_e, 1-s_2)$. Solving this equation, and using $P_e > 0$,

$$P_e = 1 - \frac{1}{2N_1}\left(\alpha_{early} - \sqrt{\alpha_{early}^2 - 4N_1\left(1 + N_1 - N_1 q_1 + \frac{\mu q(N_1 - N^* s_2)}{1+N^* s_2}\right)}\right), \quad (3)$$

with $\alpha_{early} = 1 + 2N_1 - N_1 q_1 + \frac{\mu N_1 q_1}{1+N^* s_2}$.The first order of the Taylor expansion of this expression for μ around 0 corresponds to equation (1) of the main text.

LATE MUTATION

Let us define $g_{late}(z_1,z_2)$ the generating function for one generation, starting from a virion of the initial strain when only late mutations can occur. When writing equations for the generating function, a virion of the mutant type corresponds to a term in z_2, and a virion of the initial type to a term in z_1. A virion of type 1 dies with probability $1-q_1$ without infecting a cell, leading to a term in $z_1^0 z_2^0$. Alternatively it successfully infects a cell with probability q_1, where it produces a geometric distribution of mean N_1 virions, and each of these virions is mutant with probability μ (leading to a term in z_2) or non-mutant with probability $1-\mu$ (leading to a term in z_1). That leads to:

$$g_{late}(z_1,z_2) = 1 - q_1 + \frac{q_1}{1+N_1(1-\mu z_2 - (1-\mu)z_1)}. \quad (4)$$

For the first generation after mutation, the virion may carry non-mutant proteins along with its mutant genome, so it successfully infects the next cell with probability q^* instead of q_2, thus a lineage initiated by such a virion has a survival probability $s_2 q^*/q_2$. The probability of emergence P_e starting from a virion of the initial strain is solution of $1-P_e = g_{late}(1-P_e, 1-s_2 q^*/q_2)$. Solving this equation for P_e, and using $P_e > 0$:

$$P_e = 1 - \frac{1}{2N_1(1-\mu)}\left(\alpha_{late} - \sqrt{\alpha_{late}^2 - 4N_1(1-\mu)(1+N_1(1-\mu(1-s_2 q^*/q_2))(1-q_1))}\right), \quad (5)$$

with $\alpha_{late} = 1 + N_1(1-\mu s_2 q^*/q_2) + N_1(1-\mu)(1-q_1)$. The first order of the Taylor expansion of this expression for μ around 0 corresponds to equation (3) of the main text.

A limitation of our approach is that we assume that q^* is a constant. However, if proteins can be produced from replicated genomes within the infected cell, the proportion of mutant proteins depends on the number of mutant genomes produced by a given infected cell. However, this is an issue only if μN_1 is of the order of one. Indeed, if $\mu N_1 \ll 1$, cases when a cell produces more than one mutant at the same time are very infrequent and thus can be neglected. And if $\mu N_1 \gg 1$, the standard variation in the number of mutant genomes produced by a given infected cell (which is the square root of the mean for a Poisson process) is small compared to the mean.

STAMPING MACHINE REPLICATION MECHANISM WITH MUTATIONS FOR BOTH STEPS

We have previously compared early and late mutations for a given mutation rate, in order to discuss the different effects at play. Now, we treat the general case of a stamping machine replication mechanism with a mutation rate μ_{early} for the first step and μ_{late} for the last step. We have to distinguish between virions carrying a mutant genome produced by an early mutation, which we will denote by a term in $z_{2,early}$, and virions produced after a late mutation, which we will denote by a term in $z_{2,late}$. Assuming that there is no mutation from strain 2, the generating function is:

$$g(z_1, z_{2,early}, z_{2,late}) = 1-q_1+q_1\left(\frac{\mu_{early}}{1+N^*(1-z_{2,early})} + \frac{1-\mu_{early}}{1+N_1(1-(1-\mu_{late})z_1-\mu_{late}z_{2,late})}\right). \quad (6)$$

The probability of emergence P_e starting from a virion of the initial strain is solution of $1-P_e = g(1-P_e, 1-s_2, 1-s_2q^*/q_2)$. Solving this equation, and using $P_e > 0$:

$$P_e = 1 - \frac{\alpha - \sqrt{\alpha^2 + 4N_1(1-\mu_{late})\beta}}{2N_1(1-\mu_{late})}, \quad (7)$$

with:

$$\alpha = 1+2N_1-N_1q_1-\mu_{late}N_1\left(1+s_2\frac{q^*}{q_2}-q_1\right)+\frac{\mu_{early}N_1q_1(1-\mu_{late})}{1+s_2N^*}, and \quad (8)$$

$$\beta = -1-N_1+N_1q_1+\mu_{late}N_1(1-q_1)s_2\frac{q^*}{q_2}+\mu_{early}q_1\frac{-N_1+N^*s_2+\mu_{late}N_1s_2\frac{q^*}{q_2}}{1+s_2N^*}. \quad (9)$$

The first order of the Taylor expansion of this expression for μ_{early} and μ_{late} around 0 corresponds to adding the contributions from mutations at each step, using equations (3) and (5) of the main text. Whether phenotypic mixing increases or decreases the probability of emergence depends on the phenotypic parameters and the relative mutation rates.

ANOTHER OFFSPRING DISTRIBUTION OF VIRIONS

We have assumed that infected cells produce new virions in numbers following a geometric distribution, as would be the case for an infected cell producing virions at a fixed rate per unit time which is also subject to a fixed death rate per unit time. However, as pointed out by Pearson et al. (2011), another possible scenario is that infected cells could burst and release a fixed number N virions. We explore this scenario in this section, and show that it gives results that closely parallel those shown in the main text.

First we consider the fate of a single mutant (i.e. type 2) virion. The generating function for the number of virions for one generation is now $g_{burst}(z) = 1-q_2+q_2z^{N_2}$. The survival probability is the solution of $s_{2,burst} = 1-g_{burst}(1-s_{2,burst})$, which leads to $s_{2,burst} = q_2(1-(1-s_{2,burst})^{N_2})$.

Now consider the fate of a lineage beginning with a single virion of type 1. If an early mutation occurs, N^* mutant virions are released, each with a survival probability of $s_{2,burst}$, leading to the survival probability $1-(1-$

$s_{2,burst})^{N^*}$. Following the logic in the main text, the emergence probability is then $q_1\mu(1-(1-s_{2,burst})^{N^*})/(1-q_1N_1)$, a result that is difficult to discuss in the general case. However, if $N^* = N_2$ (i.e. the phenotypic effect on N is advanced by one generation), because $s_{2,burst}/q_2 = 1-(1-s_{2,burst})^{N_2}$, the emergence probability simplifies to $\mu q_1 s_{2,burst}/(q_2(1-N_1q_1))$. In the main text, the result for a geometric offspring distribution was $q_1\mu N^* s_2/((1-N_1q_1)(1+N^*s_2))$. In the case where $N^* = N_2$, and using the result $s_2 = q_2(1-1/N_2q_2)$, the emergence probability for the geometric distribution is then $q_1\mu s_2/(q_2(1-N_1q_1))$. This is identical to the result just derived for the burst model, except for the value of the survival probability of a mutant virion.

When a late mutation occurs, we can use the same reasoning as in the main text, and again the only difference would be that $s_{2,burst}$ replaces s_2, so the discussion would be very similar.

LITERATURE CITED

Harris, T. E., 1963. The theory of branching processes. Dover Phoenix Editions, Mineola, NY.

Pearson, J. E., P. Krapivsky, and A. S. Perelson, 2011. Stochastic Theory of Early Viral Infection: Continuous versus Burst Production of Virions. PLoS Comput. Biol. 7:e1001058.